



# Systematicity, knowledge, and bias. How systematicity made clinical medicine a science

Alexander Bird<sup>1</sup> 

Received: 19 January 2016 / Accepted: 16 February 2017

© The Author(s) 2017. This article is published with open access at Springerlink.com

**Abstract** This paper shows that the history of clinical medicine in the eighteenth century supports Paul Hoyningen-Huene's thesis that there is a correlation between science and systematicity. For example, James Jurin's assessment of the safety of variolation as a protection against smallpox adopted a systematic approach to the assessment of interventions in order to eliminate sources of cognitive bias that would compromise inquiry. Clinical medicine thereby became a science. I use this confirming instance to motivate a broader hypothesis, that systematicity is a distinctive feature of science because systematicity is required by processes of knowledge generation that go beyond our everyday cognitive capacities, and these processes are required to produce knowledge of the kinds that science aims at.

**Keywords** Systematicity · Nature of science · Aim of science · Bias · Clinical medicine · Variolation

## 1 Introduction

We humans, individually and collectively, know many things. Much of what we know is acquired by the straightforward use of the cognitive capacities with which we are born and which develop in the course of everyday life. Such is everyday knowledge. In addition we have knowledge that cannot be acquired by the normal use of such capacities. This knowledge needs those capacities to be supplemented or sometimes replaced by often sophisticated means of knowledge-generation. This knowledge,

---

✉ Alexander Bird  
Alexander.Bird@bristol.ac.uk

<sup>1</sup> Department of Philosophy, University of Bristol, Cotham House, Cotham Hill, Bristol BS8 3BE, UK

which goes beyond everyday knowledge, includes scientific knowledge.<sup>1</sup> What then distinguishes scientific from other forms of knowledge, everyday knowledge in particular? Hoyningen-Huene (2013) proposes that it is the *systematicity* of science that distinguishes it from other forms of knowledge.<sup>2</sup>

The central purpose of this paper is to present one example, clinical medicine in the eighteenth century, as a confirming instance of Hoyningen-Huene's thesis.<sup>3</sup> The introduction of systematicity into medical thinking goes hand in hand with its becoming scientific.

Our ordinary cognitive capacities are limited by their liability to suffer from various kinds of bias. That liability to bias is the major source of the paucity of medical knowledge before the eighteenth century despite a long prior history of medical practice and thought. One function of systematicity is to overcome such biases. Systematicity in medicine thereby enabled it to generate knowledge that it could otherwise not have acquired.

While this is just one confirming instance of Hoyningen-Huene's thesis, it is also suggestive of a broader hypothesis, that systematicity is a mark of science because it is systematicity that enables inquiry to overcome the limitations of our ordinary cognitive capacities. And it is characteristic of science that the forms of knowledge at which it aims cannot be acquired by the straightforward application of our ordinary cognitive capacities. I conclude by discussing this hypothesis in relation to Hoyningen-Huene's work.

## 1.1 The concept of *systematicity*

Hoyningen-Huene does not offer a definition of 'systematicity'. Rather, 'systematicity' is a family resemblance concept: applications of the concept in different contexts will have some features in common but not all applications in all contexts will have the same features in common. At the most abstract level we are able to say that often systematicity will be a matter of being methodical or ordered or aiming at completeness. But not always, and even these qualities will have different forms of instantiation. Hoyningen-Huene therefore discusses nine dimensions along which science is more systematic than everyday and other kinds of non-scientific knowledge and looks at the more concrete characterizations of systematicity in each: descriptions, explanations, predictions, defence of knowledge claims, critical discourse, epistemic connectedness,

<sup>1</sup> But this non-everyday knowledge is not limited to scientific knowledge. For we can know things by sophisticated, non-everyday means without that knowledge being scientific, as when one uses infra-red goggles to identify where someone is in the dark.

<sup>2</sup> It is one of Hoyningen-Huene's merits that he focuses on this question rather than the question of demarcating science from pseudo-science. He does address the latter in the course of answering the former.

<sup>3</sup> I regard clinical medicine as the study and practice of patient care, in particular regarding diagnosis, treatment, and the maintenance of health. In this sense clinical medicine contrasts with laboratory medicine and with medical theory. (Another legitimate but different definition holds clinical medicine to be the type of medicine that developed from the late eighteenth century as a result of the birth of the hospital clinic.) In asking whether clinical medicine—'the study and practice of patient care'—is scientific, I am asking whether the study involved is scientific and whether the beliefs and knowledge informing the practice are scientific.

the ideal of completeness, the generation of new knowledge, and the representation of knowledge.

The lack of a general definition of systematicity presents no obstacle to my argument. For on any conception of ‘systematic’ the everyday ways of gathering and using evidence that were employed in much pre-modern medicine were unsystematic and the processes that supplanted them were much more systematic. For example, if I form an opinion on the basis of my memories, gathered over a long period, of what seems to be to be relevant evidence, that process will often be unsystematic. It may be unsystematic because my memories will be incomplete: not every relevant experience will be remembered. It may be unsystematic because the use of my memories will be unmethodical: which memories will be brought to bear on a hypothesis and how strongly I take them to bear on the hypothesis will depend in part on irrelevant factors, such as their being more recent or more salient for reasons unconnected with the epistemic task. If, on the other hand, my reasoning is based on data that are recorded at the relevant time, according to a prescribed plan, and which are incorporated into reasoning in a way that ensures that all relevant data are included, then my data collection and processing will be systematic because it is complete and methodical.

## 1.2 The concept of *knowledge*

The hypothesis I propose below holds that systematicity is a characteristic of science because it is necessary for scientific knowledge. And it is necessary for scientific knowledge because it is required for the reliability of scientific reasoning. The conception of knowledge I am using is one, therefore, such that for a belief to be knowledge, it must be true and generated by a reliable process. On this point Hoyningen-Huene and I differ. He says that this is a sense found in philosophy whereas he uses ‘knowledge’ ‘in the sense of a “body of ...belief that is well-established, widely held in the relevant community, not regarded as tentative or falsified” ’ (Hoyningen-Huene 2013: p. 21). He holds his sense to be the way scientists use the word. I disagree.<sup>4</sup> More importantly, I propose that the ‘philosophical’ concept of knowledge has an explanatory role in linking science and systematicity, whereas Hoyningen-Huene’s does not. That said, many of the persuasive arguments that Hoyningen-Huene mounts certainly look as if they support the view that systematicity is conducive to truth and to knowledge (in the philosophers’ sense). For example Hoyningen-Huene (2013: p. 89) argues that science is distinctively systematic in its attempts to avoid error (i.e. falsehood), including error arising from bias. That is exactly right, and this paper aims to support that proposition. Whereas, if one uses ‘knowledge’ in Hoyningen-Huene’s way, with no implication of truth, then one cannot use knowledge to link science and systematicity in the way that I do. First, the link between science and knowledge: it would be odd to think

<sup>4</sup> Certainly scientists will call ‘knowledge’ the body of belief that they accept as being non-tentative, but that’s because they think that those beliefs are true. If Hoyningen-Huene were right, then today’s scientists would be committed to agreeing that before Copernicus, astronomers genuinely *knew* that the Sun, Moon, and planets orbited the Earth. But, today’s scientists would not accept that as literally correct—they would, I am confident, say that pre-Copernican astronomers mistakenly thought they knew that the Sun orbited the Earth.

that science aims to produce an accepted body of belief—after all, that aim could be achieved by mass hypnosis. On the other hand, it makes a lot of sense to say that science aims at knowledge in the sense of (something like) reliably generated true belief. Secondly, the link between systematicity and knowledge: it might be true that being systematically produced may make claims more likely to be widely accepted. But it is more revealing to note that being systematically produced will make a claim more likely to be true (which is *why* it is more likely to be accepted).

### 1.3 Systematicity, reliability, and knowledge

Why is systematicity important to science? My claim is that systematicity is important for the reliability of the conclusions of scientific reasoning. If the inferential practices of science are not reliable then they do not produce knowledge. Unsystematic thinking is not always unreliable. Writing a shopping list is systematic; but many people have a good enough memory to be able to go to the shops without one and still know that they have bought all they need to make the intended meal. I have made no systematic study of botany, but I can still reliably distinguish an oak from a sycamore: so I do know that this tree is an oak. Thus everyday knowledge, produced without the benefit of systematicity, is possible. The kinds of proposition characteristic of science, however, are not like these everyday propositions. The processes by which we form our beliefs regarding the propositions characteristic of science are typically not reliable without systematicity. If I hit my knee against a table corner and later there is swelling and pain in my knee, then I can know without systematic investigation that the former is the cause of the latter. On the other hand, if a child is given a vaccine and some weeks later is diagnosed with autism then a causal connection between the two cannot be established in any everyday way—there is no everyday knowledge of such causal relations. A connection (whether singular or general) between a vaccine and a condition such as autism can only be established by scientific means—and that requires a systematic study of some kind. One could conduct an unsystematic study but in cases such as these, the lack of systematicity (of the relevant sort) will mean that the study is unreliable, and so no knowledge of the causal connection can be gained.

In summary my general hypothesis is this:

(H) The reasoning processes by which we come to accept the propositions characteristic of science are reliable, and so knowledge-producing, only if they are systematic.

(H) supports Hoyningen-Huene's claim that systematicity is a necessary feature of science.<sup>5</sup>

It will not be possible here to provide a comprehensive argument for (H). To do so would require careful consideration of all nine dimensions of systematicity that Hoyningen-Huene identifies, in order to show how each, in an appropriate scientific context, is necessary for the reliability of the means by which we come to believe propositions in that context. It would require arguing that there is no scientific context

<sup>5</sup> It is silent on whether systematicity is sufficient for science.

in which systematicity (in any dimension) is not necessary for the reliability of that part of science. That is too large a task for this paper. Instead I shall look at one episode in the history of medicine to show first that there is a correlation between that area of inquiry becoming increasingly scientific and the increasing systematicity of its methods. I argue that the reason for this correlation is that systematicity enable the inferences to be reliable: without a systematic approach the relevant beliefs are liable to various kinds of bias, and so are not reliably produced and so are not knowledge.

## 2 Ignorance and bias in medicine

For most of its history, the medical profession has been doing more harm than good—and a large proportion of the good that it did do was inadvertent, being the product of placebo effects.<sup>6</sup> In part that is due to the fact that medicine was unable to harness advances in wider science for the benefit of patients until the second half of the nineteenth century (Wootton 2007). Nonetheless, it is remarkable that doctors were not more knowledgeable about the causes of diseases, nor about their cures, given the great extent of medical experience. Doctors spend their lives observing patients. And the empiric school in medicine and its successors urged doctors to base their beliefs on such experience. So how is it that so many doctors observing so many patients over so many years could, in the early nineteenth century, still be prescribing antimony and mercury pills and purges, as well as bloodletting and cupping? At the same time doctors had at best hazy ideas about the correct mechanisms of disease causation while false theories (such as the miasma theory) were established doctrines.

I propose that a major part of the explanation for the failure of medicine to produce clinically useful knowledge despite the wealth of clinical experience is the fact that the latter was not systematic. Consequently, clinical medicine was for most of its history almost entirely unscientific. Only in the eighteenth century was any clinical practice subjected to and informed by the kind of systematic scrutiny that could allow it to be described as scientific. This systematicity developed further during the nineteenth century so that by the twentieth century clinical medicine was established as a science.

The use of unsystematic clinical observations to inform medical opinion was subject to a number of biases. Opinion infected by bias is likely to be erroneous; it will not be reliably formed and so will fail to constitute knowledge. As we shall see, what enabled clinical medicine to overcome such biases, and so to produce knowledge, is that it became systematic in certain ways.

The most significant of the cognitive biases to affect the progress of medicine is *confirmation bias* (Nickerson 1998). As Bacon (1620) recognized:

The human understanding when it has once adopted an opinion (either as being the received opinion or as being agreeable to itself) draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects; in order that by this great and

<sup>6</sup> Herbert Benson's (1996: p. 109) observation that 'in large measure, the history of medicine is the history of the placebo effect' is a common one.

pernicious predetermination the authority of its former conclusions may remain inviolate.

In Bacon's day, and for a long time before and a considerable time afterwards, received opinion in medicine was for the most part drawn from the works of ancient authorities, such as Galen, transmitted through the universities. The weight of such authority meant that the tendency to confirmation bias would have been particularly strong.

Another, general feature of medicine that makes it liable to confirmation bias is the inherent variability of human physiology and pathology. Few claims in clinical medicine are true and precise exceptionless generalizations. Almost every claim will have exceptions or will need to be expressed in terms of a tendency or probability. Indeed humoral medicine made this feature central to its doctrines—all disease is the result of interactions between variable states of constitution and the environment. And most medicine is imprecise—there are few clear divisions between a given kind of state and its absence (e.g. suffering from a particular condition or not). This imprecision was even greater before the advent of techniques for quantifying important variables (temperature, blood pressure, etc.).

The variability of physiology and pathology means that it is very easy to ignore contrary evidence. If every medical proposition is expected to have exceptions, then any case that does not adhere to the rule can, as Bacon noted, be put aside as a mere exception, as entirely expected, and not something that rationally threatens the truth of the underlying idea. We know that the heavy smoking nonagenarian doesn't undermine the claim that smoking causes cancer. Likewise our forebears could easily reconcile the failure of a patient to recover after a mercury purge with the underlying 'truth' that mercury purges are generally beneficial. So contrary cases were not counter-evidence, though every success could be confirming evidence.

The imprecision of medicine also invites confirmation bias. If a scientific theory is precise and makes a precise prediction that is not borne out by precise observation, then the presence of an anomaly is apparent. And so if confirmation bias is to operate in such a case it will require the conscious acknowledgment that there is at least a *prima facie* problem for the theory and some intellectual effort will be required to reconcile the theory with what was observed. If we remove the precision, then it is that much easier for confirmation bias to operate. A contrary case can easily be dismissed or ignored on the ground that it falls outside the scope of the relevant generalization (the patient didn't have the disease in question, so it is no surprise that the cure didn't work). A contrary case can even be regarded as a confirming instance, since the vagueness of the relevant states allows them to be interpreted favourably. Is the patient's pulse weak or strong (a leading diagnostic test for centuries)? It may in fact be on the weaker side of borderline, but if it varies, being sometimes weaker and sometimes stronger, then the stronger episodes may be taken to be diagnostic.

Often such cases of confirmation bias are reinforced by *measurement bias*. Measurement bias occurs when the recording of some outcome is biased, typically by some expectation of the result (the experimenter expectancy effect). In a now classic study [Rosenthal and Fode \(1963\)](#) showed that students recorded the time it took for rats to navigate a maze differently depending on their beliefs about whether the rats were bred for intelligence or not. The experiment permitted measurement bias because the

outcome being measured (whether a rat completed the maze test) was imprecise—e.g. the whiskers of a rat taking an erratic path touch the finishing line: that may be recorded as completion for what is believed to be a smart rat expected to have a fast time, but not for a normal rat. Measurement bias is a particular problem in those areas of medicine where the principal outcome is a patient's self-reported state. Has this homeopathic treatment been followed by an improvement in the patient's chronic back pain? There may be measurement bias in the patient's report (she expects to feel better, so over-emphasizes the degree to which she does feel better and under-reports the frequency or intensity of bad episodes) and in the physician's record of the patient's report.<sup>7</sup> For much of its history, the principal source of information for a doctor about a patient's state of health was of this kind—the patient's own report of how she or he felt. Doctors rarely carried out physical examinations. In the middle ages, the renaissance, and beyond, the status of the doctor as a university educated man (which is all that 'doctor' signifies) meant that medicine was an intellectual rather than a manual activity and so the appropriate source of information was verbal rather than physical. Physical contact was limited to the taking of the pulse, and observation to what could be seen in the face of the patient and examination of urine or sometimes pus. Even those observations were qualitative: while pulse watches and thermometers were available in the eighteenth century, they were used infrequently because they often conflicted with patients' reports (Nicholson 1993: p. 806). Being qualitative, the physicians' observations made without instruments were especially prone to measurement bias.

Other sources of bias played a part in eighteenth and nineteenth century medicine. Alongside confirmation bias, the *availability heuristic* can distort assessment of the evidence. We place more evidential weight on evidence that is particularly salient, for example by being especially recent in our experience, or by being what we were taught, or by being dramatic in nature (even if the drama itself is not evidentially relevant). The nature of an individual physician's experience, treating a wide variety of patients and diseases as they presented themselves, would render him liable to being misled by the availability heuristic. The evolution of medical journals in the nineteenth century was intended to broaden the horizons of practitioners, and one means by which journals achieved that was the publication of case reports. Yet such case reports, often concerning unusual or otherwise particularly interesting cases, apart from being at risk of publication bias, may also have generated bias through the availability heuristic. An effective assessment of the evidence should consider all the evidence and not weight certain cases more than others without a rationale.

### 3 Systematicity and the elimination of bias from medicine

The eighteenth century saw the beginnings of a more systematic approach to the evaluation of therapies. The first and for a long time the best example is Jurin's study of the safety of variolation. Variolation was the practice of inoculating a person, usually a child, with pus or scab from a smallpox blister in order to give them a mild case of

<sup>7</sup> The patient's measurement bias is a different phenomenon from a placebo effect, though they may often co-occur.



smallpox. That gave protection against potentially fatal naturally occurring smallpox thereafter. Variolation had been introduced to Britain in 1721 by Lady Mary Wortley Montagu, who had encountered the practice in Istanbul where she was the British ambassador's wife. It had earlier been discussed in the pages of the Royal Society's *Philosophical Transactions*. And so when a letter from the Yorkshire doctor [Nettleton \(1722\)](#) describing his success in following the new practice reached the Royal Society, its secretary, James Jurin, embarked on a study of its value.

Since it was well-known that those who survived a smallpox infection were immune thereafter, the efficacy of variolation was not itself in question. What was debatable was its safety. Was an individual who underwent variolation at a greater or lower risk of death than one who did not and so who risked a natural infection? Since it was appreciated that in both cases what is at issue is risk or chance, it was clear early on that what was needed was information about the number of those who had been variolated and the proportion of those who died, along with data on mortality from natural smallpox. Jurin was perfectly placed, intellectually and socially, to undertake this investigation, since he was both a doctor and a mathematician and was well connected through his role in the Royal Society. He used his position to solicit information from across Britain regarding those individuals who had been inoculated with smallpox. Jurin specified the details that he required (name, age, manner of inoculation, whether the inoculation successful in producing smallpox, and if so after how long, whether the patient survived or died, and so on). Jurin was also clear that he required such information concerning *all* those inoculated by those responding to his requests, thereby avoiding bias arising from the selective use of data.

Jurin published his results in a pamphlet in 1723, which he updated each year until 1727, the principal results being laid out in tables, allowing for the calculation of the risks in question.<sup>8</sup> The collation of numerical data in tables significantly reduced the liability to confirmation bias and bias through the availability heuristic. Presented in this way no case exerts more influence on the conclusion than any other; in presenting all the available data, 'unfavourable' data is as likely to be included as favourable data. Of course, that does not eliminate all bias. For there could be biases in the reports from his correspondents, the inoculators in the various parts of Britain. For this reason, Jurin took considerable pains to obtain detailed case reports rather than just the numbers. This enabled him to make uniform judgments (e.g. as to whether an inoculation had succeeded in giving the patient smallpox). He even asked for names, although these were not published, so that any controversial cases could be followed up.<sup>9</sup> As discussed above, imprecision, borderline cases, and uncertainty are especially liable to lead to bias. Jurin handled these by publishing the details of such cases so that reader could form their own judgments rather than rely on his. For example, it was not always clear whether a smallpox death following inoculation was due to that inoculation or due to a concurrent natural infection. So Jurin not only published the details of such cases but also produced tables that gave different risks depending on how many of the cases one regarded as caused by the inoculation. Jurin may have

<sup>8</sup> See, e.g., [Jurin \(1723, 1724\)](#). For discussion see [Tröhler \(2000\)](#) and [Rusnock \(2002\)](#).

<sup>9</sup> Jurin was explicit that he would preserve all the original data, so that particular cases could be followed up, e.g. to see whether someone who had natural smallpox had previously been inoculated.



been especially keen to be transparent in order to avoid imputation of bias because he was acutely aware of the effect of confirmation bias and the availability heuristic on public opinion when it came to this controversial practice. As one correspondent put it, 'If anything goes amiss or seems to do so, the world presently sings of it with all the Aggravation imaginable, but on the other hand many Successfull Experiments are I believe buried in silence.'<sup>10</sup> With this in mind, Jurin became very active in following up cases that appeared to question the safety of variolation.

The eighteenth century saw further developments that made the collecting of evidence in clinical medicine more systematic and so less liable to bias and other forms of error that would undermine the reliability of inferences drawn from that evidence. Imprecise qualitative reports were increasingly replaced by numerical data. For example, because smallpox could scar or even blind those who survived a bad case, it was important that inoculation gave only mild cases of the disease (but genuine cases nonetheless). Many doctors held that the risk of a severe eruption could be reduced by the use of laxatives. Sir William [Watson \(1768\)](#) undertook a series of experiments to test this hypothesis and also to test different types of material used in the inoculation ([Boylston 2008](#)). But what counts as a 'severe' eruption? If the experiment were to be repeated elsewhere, how could one know that other researchers would regard the same cases as equally severe by their standards? Watson eschewed qualitative assessment in favour of counting the number of smallpox pustules, thereby ensuring comparability of cases while also reducing the room for measurement/experimenter bias.

Another eighteenth century innovation that promoted systematicity and thereby the reliability of data and of the inferences drawn from them was the introduction of the blank form. This was particularly helpful where research required input from multiple collaborators. In his collaborative research, Jurin requested detailed case studies and had to follow up many reports for important missing information. He and other researchers appreciated that they could ensure that they received the data they required while also reducing the labour of their correspondents by supplying those correspondents with blank forms or recommending the use of specific structured tables for recording observations. In another project, Jurin resurrected and adapted a much earlier (1667) proposal of Hooke's for the collection of meteorological data. Jurin's project sought to collate data from across Europe and North America, with the aim of linking the weather and medicine. Correspondents were given detailed instructions about the data they were to collect and how it was to be presented, along with a specimen form that they could copy ([Rusnock 2002](#): pp. 111–114). Half a century later a very similar project was undertaken within France and its colonies by Félix Vicq d'Azyr and the Société Royale de Médecine ([Rusnock 2002](#): pp. 117–118). Vicq d'Azyr was able to use the bureaucracy of the Société and the French state to ensure a greater degree of engagement than Jurin had been able to achieve, and was aided by the use of a form that had dates and times printed in advance, making it difficult for participants to avoid making and recording the required observations. In 1731 Francis Clifton had proposed a standard tabular form for the recording of medical case histories. While his innovation was not widely adopted in the medical profession, the idea that standard forms and

<sup>10</sup> Letter from Edward Edlin to Jurin, 1726, quoted in [Rusnock \(2002](#): p. 64).

tables would promote the uniform, comparable, and unbiased collection of data was nonetheless established. In his ‘An Attempt to improve the Evidence of Medicine’ George Fordyce (1793) promoted the use of a standard table for the recording of patient histories with benefits both to patients and to science. Johann George Christoph Siebold did likewise at the Julius Hospital in Würzburg (Hess and Mendelsohn 2010: p. 292). Pierre Charles Alexandre Louis noted Fordyce’s proposal in his influential work on bloodletting (Louis 1835). Louis remarked that whereas Fordyce had the general practitioner in mind, which raised difficulties in the promotion and use of his scheme, Louis himself had the advantage of working in a hospital where it was easier to enforce such uniformity. Louis’s own work involved the systematic collection and analysis of data on bloodletting and its relation to mortality from pneumonitis—he showed that bloodletting early after diagnosis did not have any advantage over bloodletting later in the treatment. The use of blank forms and standardized processes of collecting data can be seen as components of a broader process of the systematization of record taking, collation, and publication, originating with books of unconnected patient histories and leading to carefully constructed case series, that, argue Hess and Mendelsohn (2010: p. 287), ‘became a basic operation of medical knowing’.

If one wants to know whether an intervention (e.g. a treatment) caused a certain outcome (e.g. recovery) in a particular patient, then one will typically want to know whether the outcome would have occurred or not without the intervention. If one’s background belief is that this intervention is causally effective and one is subject to confirmation bias or the availability heuristic, then one will tend to focus on those cases of the intervention that are followed by the expected outcome, and perhaps on those cases where the lack of intervention was followed by the lack of the outcome (no treatment followed by death). But such biases will tend to hide the cases where a lack of intervention was nonetheless followed by the outcome in question, e.g. a patient recovered despite the lack of treatment. It may appear that the connection is causal but one cannot know that it is—confirmation bias is a close cousin of the fallacy *post hoc ergo propter hoc*. Yet knowing the rate of such spontaneous recovery is crucial to knowing whether the treatment is indeed causally effective. So knowledge of causal relations from observations of patient outcomes requires data not only on those who were treated but also on those who were not treated but might have been. Furthermore, that causal knowledge requires all the data or at least an unbiased sample, not the biased sample produced by confirmation bias or the availability heuristic.<sup>11</sup> This is what is achieved by the collection of data that is systematic in that data on all patients is used—or data on all patients selected by a process fixed in advance so as not to be biased towards any particular outcome.

Instances where the intervention is absent are *controls*. While the significance of controls in assessing causal claims was recognized beforehand, the eighteenth century saw the first systematic use of controls for the assessing causal claims. Lind’s 1753 small scale study of different treatments for scurvy is a well-known, indeed iconic example. In 1816 Alexander Lesassier Hamilton wrote of an experiment that he and colleagues performed during the Peninsula War, in which soldiers were assigned to

<sup>11</sup> Unbiased information on those who were not treated is *necessary* for the production of causal knowledge from observational studies, but it is not *sufficient*.

one of three surgeons, one of whom used the lancet (for bloodletting) while the other two did not (Milne and Chalmers 2015). He recorded ‘Neither Mr Anderson nor I ever once employed the lancet. He lost two, I four cases; whilst out of the other third thirty-five patients died.’ Significantly, Hamilton and his colleagues adopted systematic procedures in order to eliminate bias: the soldiers were assigned to the three surgeons by alternation, thus avoiding allocation bias (whereby soldiers with differing severity of illness might have been differentially allocated to the three groups.) The soldiers were received into the trial ‘indiscriminately’—reinforcing the avoidance of allocation bias (allocation bias could have occurred with alternation, if, for example, knowledge that the next soldier to be admitted would go to Mr Anderson would have affected whether he was admitted to the trial at all). The soldiers ‘were attended as nearly as possible with the same care and accommodated with the same comforts’, to avoid bias arising from an imbalance of the groups in the trial. Guy (1860) records a trial carried out by Dr Thomas Balfour on boys at the Royal Military Asylum regarding the use of belladonna as a prophylactic against scarlet fever. 151 boys were divided into two groups ‘taking them alternately from the list, to prevent the imputation of selection’. 76 were thereby given belladonna and 75 were not; two in each group contracted scarlatina. The systematic nature of the allocation to eliminate bias is again to be noted. Also significant is Balfour’s comment about the liability to bias in an unsystematic study, ‘The numbers are too small to justify deductions as to the prophylactic power of belladonna, but the observation is good, because it shows how apt we are to be misled by imperfect observation. Had I given the remedy to all the boys, I should probably have attributed to it the cessation of the epidemic’ (quoted in Guy 1860: p. 554).

#### 4 Systematicity, knowledge, and the nature of science

Hoyningen-Huene’s (2013) central thesis is:

(S) Scientific knowledge differs from other kinds of knowledge, in particular from everyday knowledge, primarily by being more systematic (2013: p. 14).

(S) suggests two further theses concerning the genesis and development of a new science:

(G) The newborn science is more systematic than the knowledge practice from which it emerged (2013: p. 180),

and

(D) In the further development of a scientific discipline or field, its overall systematicity will increase (2013: p. 183).

In the preceding section I have presented evidence in favour of (G). There was little or nothing that could be called clinical science before the eighteenth century whereas Jurin’s work on variolation and similar work by his contemporaries and successors shows the hallmarks of modern, scientific clinical medicine in a nascent but nonetheless clear form. The discussion revealed the importance of systematicity in these developments whereby clinical medicine became scientific. A systematic approach to the

investigation of certain procedures enabled doctors to begin to know what was safe and effective—this was knowledge they could not have acquired any other way. That is because the systematic approach allows for the elimination (or at least reduction) of the influence of bias on results. A doctor's own experience is often insufficient to produce knowledge that one treatment leads to a satisfactory outcome in a higher proportion of cases than some other treatment. The safety of variolation was an issue of precisely this sort: was the chance of death lessened or increased by this procedure? The number of inoculations carried out by most inoculators in the early days of variolation was too small for any individual doctor to know that it decreased the overall risk of mortality. The health of children and the risk of their death are emotionally charged matters; there were also religious currents in the debate over variolation. In such circumstances the perceptions of doctors and the public were very likely to be clouded by confirmation bias and the availability heuristic. By collecting and analysing the data in a systematic manner Jurin overcame such biases and was able to know that a smallpox inoculation carried a lower chance of death than the risk run by the uninoculated of dying from natural smallpox.

The brief history also provides some support for (D)—developments later in the eighteenth century and into the nineteenth century showed further additions to systematicity in clinical medicine (such as the introduction of structured blank forms). And, yielding further evidence in support of (D), these have continued to the present day. The Evidence-Based Medicine (EBM) movement was instituted, in the minds of its proponents, to combat remaining sources of bias in the evaluation of treatments. In saying this I do not intend to endorse all claims made on behalf of the EBM movement. Rather I am pointing out that the movement *sees itself* as one that is pushing for increased systematicity in the evaluation of treatments (and in medical education) in order to reduce bias and so increase knowledge. I note, for example, that one key set of the products of EBM, clinical guidelines, are defined as 'systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances' (Field and Lohr 1990). There are debates about whether EBM in fact achieves this (see Howick 2011 for discussion). Furthermore there are important debates about whether it achieves this at too great a cost: epistemically (by downplaying other routes to genuine knowledge) as well as ethically (Worrall 2006). My own view (Bird 2010) is that EBM has made a valuable contribution to medicine's ability to produce genuine knowledge, but this should not obscure the fact that in some circumstances at least there are ways of producing knowledge that are not based on randomized controlled trials. For example, observational studies cannot systematically eliminate bias. Nonetheless, the results of some observational studies cannot be explained away just by the biases to which they might be liable. So in such a case the best explanation for the size of some observed effect is that the treatment under test is indeed effective. In extreme cases everyday means of cognition can suffice to produce knowledge of scientific significance. No systematization was needed to see that the doses of penicillin given to Albert Alexander caused the almost miraculous (but temporary) remission of the infection that eventually killed him, such was the unprecedented degree of that remission. So EBM needs to accept the principle that sometimes knowledge can be gained from evidence of a type low down in its hierarchy. Does this admission undermine my case that systematicity is neces-

sary for science to produce knowledge? Hardly. Such cases are very rare in medicine. Much more frequently we are looking for rather more marginal gains for a population ('on average patients taking this drug will live six months longer'). Such gains cannot be detected without the systematic approach of a clinical trial. Moreover, there will be other clinically important aspects of a treatment that may only be revealed by a clinical trial—e.g. concerning safety and side-effects or long-term treatment outcomes. It might have been possible without a clinical trial to tell that streptomycin is active against *Mycobacterium tuberculosis* in TB patients. But without that trial and its systematic recording of data researchers would not have discovered that its effectiveness reduces over time—the first observed instance of antibiotic resistance (Medical Research Council 1948; Crofton 2006).

At the peak of the EBM hierarchy are systematic reviews. As Hoyningen-Huene (2013: p. 102) notes, systematic reviews gather and analyse results from multiple studies in a systematic way. The aim of a systematic review is to avoid the bias of cherry-picking favourable studies and ignoring unfavourable ones. A systematic review is so-called because in order to avoid bias it has explicit criteria for searching for, including, and evaluating studies. The AllTrials campaign aims to eliminate publication bias, calling for the systematic registration of all clinical trials, thereby enabling authors of systematic reviews to know whether they are including all the evidence.<sup>12</sup> The use of standard forms pioneered by Clifton, Siebold, and others was an early step in a long history of increasing standardization of many aspects of medicine (Timmermans and Berg 2010), frequently in aid of the epistemic benefits of systematicity (though standardization can serve many other purposes besides, including bureaucratic, political, and economic purposes).

For this reason and for the many he presents in his Hoyningen-Huene (2013), I find that Hoyningen-Huene makes a convincing case that science is correlated with systematicity. This in turn invites the question: why is (S) true? Is it simply that 'science' is the term we give to systematic knowledge and belief? Or is there a more profound reason for the connection? The hypothesis I proposed above is:

(H) The reasoning processes by which we come to accept the propositions characteristic of science are reliable, and so knowledge-producing, only if they are systematic.

The proposal is that the propositions of science are typically such that our attempts to know them will not be reliable unless our means incorporate some relevant kind of systematicity.

Innate cognitive capacities and cognitive capacities acquired in the course of everyday life include the senses, memory, and the ability of think fairly rationally about matters of modest complexity. These are perfectly capable of delivering everyday knowledge, since they are often reliable for those purposes. However, for many propositions these means are insufficient. These propositions include the propositions of science. If (H) is right, then the reason that the propositions of science cannot be known using our everyday cognitive capacities is that the latter are insufficiently systematic

<sup>12</sup> <http://www.alltrials.net> last accessed 19 January 2016.

to be reliable ways of ascertaining the truth of such propositions. The systematicity of science overcomes those limitations. With the right kind of systematicity, our reasoning processes can be reliable enough to know those propositions; without it, they cannot be. The following illustrates this idea.

I can know, using just memory of my experiences that where I live the weather is colder on average this week than last week. However, it is just not possible using ordinary cognitive capacities to come to know that on average temperatures in this region are warmer than they were forty years ago. My memories are not sufficiently fine-grained; they are incomplete; and they are liable to bias. I may think that the weather is now warmer, because I can clearly remember the snowy winters of my childhood whereas several recent winters have been notable for the early arrival of daffodils in my garden. Clearly, beliefs formed in the latter way will be hugely unreliable, not least because of the biasing effects of the availability heuristic: what I remember of past winters is not determined by their representativeness of the prevailing climate but instead by their relevance to my prevailing interests. Knowledge of the effectiveness and safety of a medical procedure such as varicella is just like this. Knowledge of such changes and effects can only be acquired by the systematic collection of data and its systematic analysis. Without systematicity there is no knowledge, not even unsystematic knowledge.

So in this case knowledge of a kind that it is within the aim of science to acquire—knowledge of climatic changes over long periods—cannot be acquired by ordinary cognitive capacities, because those are limited or deficient in certain respects that renders them unreliable means of finding the truth of propositions concerning this subject matter. Systematicity in the collection and analysis of data remedies those deficiencies and overcomes those limitations; it renders the reasoning reliable and so able to produce knowledge.

In this paper I have concentrated on deficiencies arising from cognitive biases and have also mentioned the limitations of memory. These deficiencies mean that our ordinary cognitive capacities, although reliable enough for many everyday purposes, need supplementation or replacement by a systematic approach in order to generate genuine knowledge in the more complex cases characteristic of science. Still, these are just examples of where systematicity enables a cognitive process to be reliable. In *these* cases we find systematicity in science because of its necessary role in generating genuine knowledge (as opposed to unreliably formed belief). But that does not show that systematicity always or even typically is associated with science. For that one would have to show that the propositions of science are typically such that they cannot be known by methods that are not systematic at all, i.e. that, normally, any humanly usable method that allows a proposition of science to be known must be systematic in a relevant way (i.e. in a way that our everyday methods of knowing are normally not systematic). This claim is plausible, insofar as one can multiply examples. Caroline and William Herschel could not have come to know facts about new nebulae and comets without Flamsteed's systematic collection of data on stars and planets in his catalogue (which they first had to correct and improve). Mendeleev's systematization of the elements in his periodic table enabled knowledge of atomic structure as well as the discovery of new elements. And so on. However, it may remain the case that there are classes of scientific proposition that can be known to be true by methods

that are not systematic in relevant respects. How can (H) be defended against such a possibility?

One approach would be to appropriate and build on Hoyningen-Huene's own arguments. Assume that we accept his detailed arguments for the conclusion that systematicity in one of its nine dimensions is to be found in any piece of science. We may then investigate those dimensions to argue that in each case that species of systematicity enables reliable methods of inquiry into the relevant scientific propositions, in that these methods would not be possible or would not be reliable without that systematicity. It is not possible for reasons of space to carry out this task in this paper. Nonetheless we can see how this investigation might proceed. The importance of systematicity for knowledge is explicitly mentioned in three of the nine dimensions. And it is implicit in the remainder. For example, critical discourse is another dimension of systematicity that is essential to reliable knowledge-generation, because it enables the elimination of false theories, without which positive theoretical claims cannot achieve knowledge (Bird 2007). In another example, classification is a component of the description dimension that is systematic in science. I have just mentioned the importance of the systematicity in Flamsteed's catalogue and Mendeleev's periodic table for the generation of new knowledge; likewise one can readily see that a systematic approach to classification, as exemplified in Linnaeus's binomial system and William Farr's nosology, was essential to progress in biology and epidemiology respectively—without it certain species of knowledge in those fields could not be had.

## 5 Conclusion

Above I conjectured this hypothesis:

(H) The reasoning processes by which we come to accept the propositions characteristic of science are reliable, and so knowledge-producing, only if they are systematic.

which I related to Hoyningen-Huene's:

(S) Scientific knowledge differs from other kinds of knowledge, in particular from everyday knowledge, primarily by being more systematic.

The two are mutually reinforcing in that (H) supports the component of (S) that says that systematicity is necessary for scientific knowledge. At the same time, I have suggested that Hoyningen-Huene's arguments for the correlation in (S) can be supplemented to show that in each case where Hoyningen-Huene finds that correlation, we can see that the systematicity of science enables the methods of science to be reliable, so that without that systematicity, the relevant scientific propositions could not be known. Completing that argument for (H) would require an extended (and systematic) argument that here I have only sketched. Nonetheless, I have given a concrete instance of the general case. Before the eighteenth century clinical medicine was not scientific. By the end of the century, if not yet a fully-fledged science, it had at least become scientific in certain respects. And this is because it became systematic in



respects that allowed it to overcome the limitations of our everyday means of belief-formation, most notably our susceptibility to cognitive biases.

**Acknowledgements** I am grateful to the two anonymous referees for very helpful suggestions and comments.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bacon, F. (1620). *Novum Organum*. In Vickers, B. (Ed.), *Francis Bacon the major works*. Oxford: Oxford University Press. Oxford World's Classics edition 2008.
- Benson, H. (1996). *Timeless healing. The power and biology of belief*. New York: Fireside.
- Bird, A. (2007). Inference to the only explanation. *Philosophy and Phenomenological Research*, 74, 424–432.
- Bird, A. (2010). The epistemological function of Hill's criteria. *Preventive Medicine*, 53, 242–245.
- Boylston, A. W. (2008). William Watson's use of controlled clinical experiments in 1768. *JLL Bulletin: Commentaries on the History of Treatment Evaluation*. <http://www.jameslindlibrary.org/articles/william-watsons-use-of-controlled-clinical-experiments-in-1767/>.
- Crofton, J. (2006). The MRC randomized trial of streptomycin and its legacy: A view from the clinical front line. *Journal of the Royal Society of Medicine*, 99, 531–534.
- Field, M. J., & Lohr, K. N. (Eds.). (1990). *Clinical practice guidelines: Directions for a new program*. Washington DC: National Academies Press.
- Fordyce, G. (1793). An attempt to improve the evidence of medicine. In *Transactions of a Society for the Improvement of Medical and Chirurgical Knowledge*. London: J. Johnson.
- Guy, W. (1860). Croonian lectures, part V. *British Medical Journal*, 86, 553–555.
- Hess, V., & Mendelsohn, A. (2010). Case and series: Medical knowledge and paper technology, 1600–1900. *History of Science*, 48, 287–314.
- Howick, J. (2011). *The philosophy of evidence-based medicine*. Chichester: Wiley.
- Hoyningen-Huene, P. (2013). *Systematicity: The nature of science*. Oxford: Oxford University Press.
- Jurin, J. (1723). *A letter to the learned Caleb Cotesworth, M. D. Fellow of the Royal Society, of the College of Physicians, and Physician to St. Thomas's Hospital. Containing a comparison between the mortality of the natural small pox, and that given by inoculation*. London: W. and J. Innys.
- Jurin, J. (1724). *An account of the success of inoculating the small-pox in Great Britain. With a comparison between the miscarriages in that practice, and the mortality of the natural small-pox*. London: J. Peele.
- Lind, J. (1753). *A Treatise of the Scurvy in Three Parts. Containing an inquiry into the nature, causes and cure of that disease, together with a critical and chronological view of what has been published on the subject*. London: Miller.
- Louis, P. C. A. (1835). *Recherches sur les effets de la saignée dans quelques maladies inflammatoires et sur l'action de l'émétique et des vésicatoires dans la pneumonie*. Paris: Baillière.
- Medical Research Council. (1948). Streptomycin treatment of pulmonary tuberculosis. *British Medical Journal*, 4582, 769–782.
- Milne, I., & Chalmers, I. (2015). Alexander Lesassier Hamilton's 1816 report of a controlled trial of bloodletting. *Journal of the Royal Society of Medicine*, 108, 68–70.
- Nettleton, T. (1722). Part of a letter from Dr. Nettleton, physician at Halifax, to Dr. Jurin, R. S. Sec concerning the inoculation of the small pox, and the mortality of that distemper in the natural way. *Philosophical Transactions of the Royal Society of London (1722–1723)* 32, (pp. 209–212).
- Nicholson, M. (1993). The art of diagnosis: Medicine and the five senses. In W. F. Bynum & R. Porter (Eds.), *Companion encyclopedia of the history of medicine* (Vol. 2, pp. 801–825). London: Routledge.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175–220.

- Rosenthal, R., & Fode, K. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, 8, 183–189.
- Rusnock, A. A. (2002). *Vital accounts. Quantifying health and population in eighteenth-century England and France*. Cambridge: Cambridge University Press.
- Timmermans, S., & Berg, M. (2010). *The gold standard*. Philadelphia: Temple University Press.
- Tröhler, U. (2000). *To improve the evidence of medicine The 18th Century British origins of a critical approach*. Edinburgh: Royal College of Physicians of Edinburgh.
- Watson, W. (1768). *An account of a series of experiments, instituted with a view of ascertaining the most successful method of inoculating the smallpox*. London: J. Nourse.
- Wootton, D. (2007). *Bad medicine: Doctors doing harm since Hippocrates*. New York: Oxford University Press.
- Worrall, J. (2006). Why randomize? Evidence and ethics in clinical trials. In W. J. Gonzalez & J. Alcolea (Eds.), *Contemporary perspectives in philosophy and methodology of science*. Netbiblio: Isla Cristina.